# Molecular analysis of a U3 RNA gene locus in tomato: transcription signals, the coding region, expression in transgenic tobacco plants and tandemly repeated pseudogenes

Tamás Kiss and Ferenc Solymosy
Institute of Plant Physiology, Biological Research Center, Hungarian Academy of Sciences, H-6701 Szeged, PO Box 521, Hungary

Received March 9, 1990; Accepted March 27, 1990 EMBL accession no. X14411

## ABSTRACT

**By screening a tomato genomic library with a tomato U3 RNA probe, we detected a U3 genomic locus whose coding region was determined by primer extension (5′ end) and direct RNA sequencing of purified U3 RNA from tomato (3′ end). Tomato U3 RNA is 216 nucleotides long, contains all the four evolutionarily highly conserved sequence blocks (Boxes A to D), has at its 5′ end a cap not precipitable with anti-$m_3G$ antibodies and can be folded into a peculiar secondary structure with two stem-loops at its 5′ end. A tagged derivative of the U3 gene was faithfully expressed in transgenic tobacco plants. In the 5′ flanking region both plant-specific UsnRNA transcription signals [the TATA-like sequence and the upstream sequence element (USE)] were present, but were positioned closer to each other and also to the cap site in the U3 gene than in the genes for the plant spliceosomal UsnRNAs studied so far. The 3′ flanking region of the tomato U3 gene lacked the consensus sequence of the putative termination signal established for the plant spliceosomal UsnRNA genes and contained a pyrimidine-rich tract (R1) followed by four tandemly repeated U3 pseudogenes (U3.1 ps to U3.4 ps) flanked by slightly altered forms (R2 to R5) of R1 and most probably generated by DNA-mediated events. Our results are in line with the conjecture that the enzyme transcribing the tomato U3 gene has different structural requirements for transcriptional activity than the enzyme transcribing plant U1, U2 and U5 genes.**

## INTRODUCTION

Nuclei from all eukaryotes studied so far contain a set of six major, capped, uridylate-rich small nuclear RNAs [UsnRNAs: U1 to U6 RNAs, cf. (1)] which, in the form of ribonucleoprotein particles (UsnRNPs), are involved in nuclear RNA processing events. Five of them (U1, U2, U4–U6 RNAs), the nucleoplasmic, so-called spliceosomal UsnRNAs (2), play a fundamental role in the splicing of pre-mRNA (2,3), whereas the nucleolar U3 RNA is involved in the processing of pre-rRNA (4–9). For a better structural and functional characterization of the individual UsnRNA species, a considerable effort has been devoted in recent years to investigating the evolutionary conservation of both the structure of these molecules and the structure and mode of expression of their genes. In this approach comparative studies of metazoan, yeast and plant systems proved to be particularly useful with the spliceosomal UsnRNAs [cf. (2)], but have failed to yield phylogenetic support to a generalized view of the structure, function and genomic set-up of U3 RNA.

With the spliceosomal UsnRNAs, including those from plants (10–14), we have a fairly comprehensive picture of the phylogenetic conservation of their primary and secondary structure [cf. (2)]. With U3 RNA, however, our information about analogous aspects of the evolutionary conservation of molecular structure is incomplete: (i) The canonical 5′ end $m_3^{2,2,7}G$ ($m_3G$) cap structure, characteristic of the spliceosomal UsnRNAs (except U6 RNA) was detected in the U3 RNAs from all organisms tested so far (15–17), except pea plants in which no RNA with the electrophoretic mobility of U3 RNA could be discerned upon precipitation of a nuclear RNA extract with anticap antibodies (10). (ii) Both yeast (16,18) and metazoan (17,19) U3 RNAs contain four evolutionarily conserved sequence blocks (Boxes A to D). Owing to the availability of only a partial (3′ end) nucleotide sequence of a plant U3 RNA (20), it is not known whether or not plant U3 RNA also contains all four of these sequence blocks. (iii) Several secondary structure models have been proposed for U3 RNA from various organisms (16,17,19,21), but none of them exhibited a degree of evolutionary conservation similar to that arrived at by analyzing individual spliceosomal UsnRNAs (2). Recently, it has been suggested (16) that the 5′ one-third of lower eukaryotic U3 RNAs folds into two stem-loops whereas the corresponding region of higher eukaryotic U3 RNAs forms a single stem-loop. Owing to the lack of the nucleotide sequence of the 5′ region of any plant U3 RNA (20), the above assumption is apparently in need of additional phylogenetic testing.

As far as the transcription of the spliceosomal UsnRNA genes is concerned, a very clear distinction between the promoters of vertebrate versus plant U1, U2 and U5 snRNA genes could be made. In vertebrates, these genes transcribed by RNA polymerase

II [cf. (22)], have a proximal sequence element (PSE) at about −50 and a distal sequence element (DSE) at about −200 or −220 in their 5′ flanking region [cf. (23)]. PSE is a TATA-box equivalent (with no sequence resemblance to the canonical TATA-box whatsoever) and is responsible for the correct initiation of transcription. DSE encompassing the octamer motif ATGCAAAT and an Sp1 binding site is an enhancer-like element. In their 3′ flanking region at about +10 or +16 these genes have a loosely conserved sequence, the '3′ Box', which has been shown in some cases to be required for proper 3′ end formation [cf. (22,23)]. In plants, the genes for U1 (24−26), U2 (27−29) and U5 (30) RNAs do have in their 5′ flanking region the canonical TATA Box centered around nt position −30, which functions as the vertebrate PSE (31) and they have between nt positions −70 and −80 a so-called upstream sequence element (USE) with no resemblance in sequence to the vertebrate DSE. The 3′ flanking region invariably starts with a CA dinucleotide residue and is followed about 4 to 9 nt residues further downstream by a 3′ consensus sequence different from the vertebrate '3′ Box'. In mammals, U3 RNA seems to be transcribed by RNA polymerase II similarly to the spliceosomal U1, U2, U4 and U5 RNAs [cf. (22)], and the transcription signals in its genes are also similar to those of the above spliceosomal UsnRNAs. There is, however, a U3-specific motif in the mammalian genes that encode U3 RNA (32,33). This is, in addition to nucleolar location and separate function, another attribute of U3 RNA which sets it apart from the spliceosomal UsnRNAs. Plant U3 RNAs have not been characterized so far at the genomic level.

In an attempt to clear up the above uncertainties concerning the extent of phylogenetic conservation of some structural elements of U3 RNA and the transcription of its gene(s) we undertook a molecular analysis of a tomato *bona fide* U3 RNA gene and its transcript. In this paper we (i) establish for the first time the complete primary structure of U3 RNA from plant material and propose a model for its secondary structure, (ii) analyze the unusual organization and structure of transcription signals in the 5′ and 3′ flanking regions of the tomato U3 gene, and (iii) report for the first time the occurrence in a plant genome of UsnRNA pseudogenes probably generated by DNA-mediated events.

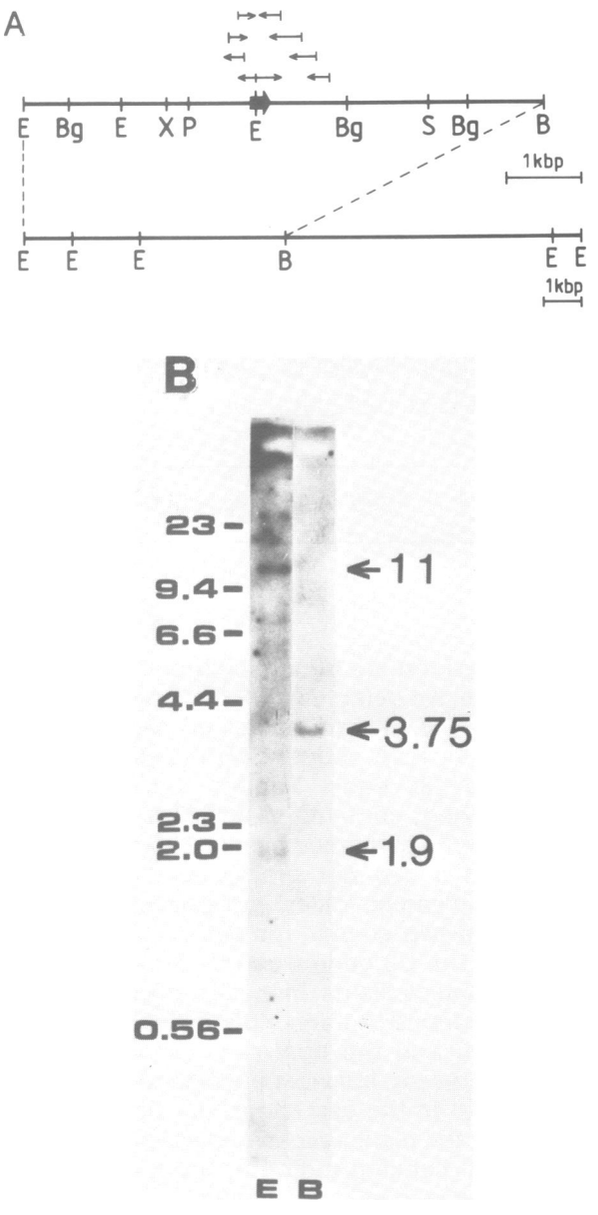


**Fig. 1.** Organization of the tomato U3 genomic locus. (**A**) Restriction map. Nine recombinant lambda phage were isolated and these carried DNA fragments which contained an identical locus of the tomato genome (below). An expanded view of the sub-cloned fragment is shown above the map of the isolated U3 locus. The U3 coding region is illustrated by a bold arrow pointing in the direction of transcription. The sequencing strategy is indicated by fine arrows at the top. Restriction enzyme abbreviations: B, *BamHI*; Bg, *BglII*; E, *EcoRI*; P, *PstI*; S, *SacI*; X, *XbaI*. (**B**) Genomic blot analysis of tomato nuclear DNA using an *in vitro* synthesized antisense U3 RNA probe (Probe A, see Materials and Methods). The DNA was cut with *EcoRI* (*Lane E*) and *BglII* (*Lane B*). Molecular size markers and the estimated lengths (kbp) of the hybridizing fragments are indicated on the left and on the right, respectively.

## MATERIALS AND METHODS

All methods, unless stated otherwise, were taken from (34).

### Isolation and fractionation of nuclear and total RNA from plants

Isolation of nuclei from tomato, tobacco and broad bean leaves, extraction of nuclear and plant total RNA, fractionation of low-molecular-weight RNAs on denaturing polyacrylamide gels and 3′ end-labeling of RNA with [$^{32}$P]pCp by T4 RNA ligase were done as described earlier [cf. (20)].

### Immunoprecipitation of RNAs with anti-$m_3G$ IgG

(kindly provided by R. Lührmann) was done as described in (10).

### Screening of a tomato genomic library

A tomato genomic library constructed in Charon 4 vector by inserting *Lycopersicon esculentum* Mill DNA partially digested with *EcoRI* was kindly provided by R.W. Breidenbach. Approximately 200,000 recombinant phage plaques were screened by the in situ hybridization procedure of Benton and Davis (35), using 3′end-labeled tomato U3 RNA as a probe. The filters were hybridized for 16 h at 42°C in 50% formamide, 5×SSC, 0.1% SDS, 1 mM EDTA, 10 mM Tris/HCl, pH 7.5, and 1×Denhardt's reagent [0.02% (w/v) each of bovine serum albumin, polyvinylpyrrolidone and Ficoll] and were washed for 2×20 min at room temperature in 2×SSC. To prevent the isolation of ribosomal DNA clones, purified tomato total RNA was added to the hybridization mixture (200 μg/ml). After three

rounds of rescreening, nine recombinant phage were isolated which hybridized strongly to tomato U3 RNA and DNA was isolated from them according to Helms et al.(36).

### Subcloning and DNA sequencing

The 1.2 kbp *XbaI/EcoRI* fragment (see Fig. 1A for restriction map) and the 1.2 kbp *EcoRI/BglII* fragment carrying the 5′ and 3′ regions of the U3 gene, respectively, were cloned into the *XbaI/EcoRI* and *EcoRI/BamHI* sites of pBluescribe M13 (+/−) vectors (Stratagene). The 2.4 kb *XbaI/BglII* fragment carrying the entire U3 gene was inserted into the *XbaI/BamHI* sites of pBluescribe M13 (+/−) vectors. To facilitate DNA sequence analysis, using the above constructs, overlapping sets of deletions were created by unidirectional digestion with Exonuclease III (37). Single-stranded DNA was isolated using R408 helper phage (38). The DNA was sequenced essentially according to Sanger et al. (39).

### Preparation of RNA and DNA probes

*Probe A.* High-specific-activity, antisense RNA probe used for genomic blot analysis of tomato nuclear DNA and for testing the in vivo expression of the tomato U3 gene in transgenic tobacco plants was synthesized *in vitro* with T7 polymerase as described (40) for SP6 polymerase, using [$\alpha^{32}$P] UTP and a *HindIII*-linearized DNA template which carried, in addition to the coding region of U3 RNA, 37 nt upstream and 42 nt downstream sequences. The above construct was generated by bidirectional Exonuclease III treatment of the 2.4 kbp *XbaI/BglII* fragment (see Fig. 1A). *Probe B.* Uniformly labeled ssDNA complementary to the inserted marker sequence used in the analysis of the in vivo expression of the tomato U3 gene in transgenic tobacco plants was synthesized by Klenow polymerase using single-stranded pBluescribe M13 (−) vector as a template and the M13 universal sequencing primer. The resulting double-stranded plasmid was linearized by digestion with *BamHI* and the ssDNA probe was purified on a 6% sequencing gel.

### Southern analysis

Tomato nuclear DNA was isolated as described earlier (41). About 8−10 mg of DNA were digested with *EcoRI* or *BglII* (see Fig. 1A) and separated by electrophoresis through 0.8% agarose gel.

### RNA sequence analysis

Enzymatic sequencing of 3′ end-labeled tomato U3 RNA was performed as described in (14).

### Determination of the 5′end of tomato U3 RNA

The 5′ end of the tomato U3 RNA was determined by primer extension assay. The coding region of tomato U3 gene contains an *RsaI* site (position 38) and an *EcoRI* site (position 53). This internal *RsaI/EcoRI* fragment was isolated and labeled at its 5′ ends using T4 polynucleotide kinase. The antisense strand (19 nt) was separated from the sense strand (15 nt) by polyacrylamyde gel electrophoresis and was used as a primer both for the reverse transcription and for the DNA sequencing reactions. The reverse transcription reaction was carried out in the presence of 2 μg of tomato nuclear RNA and 1 pmol purified primer (5000 c.p.m.) in 50 mM Tris/HCl, pH 8.3, 10 mM $MgCl_2$, 20 mM 2-ME, 100 μM dTTP, dGTP, dCTP each and 2 μCi [$^{32}$P] dATP (400 Ci/mmol) with 5 U AMV reverse transcriptase at 42°C for 30 min. Both the reverse transcription and the DNA sequencing reaction products were ethanol-precipitated and analyzed on a 6% sequencing gel.

### Determination of the possible secondary structure of tomato U3 RNA

This was done by taking the free energy values from (43) and the folding program of Zucker and Stiegler (44) without forced stacking.

### Analysis of tomato U3 gene expression

For the construction of an insertion mutant of the tomato U3 gene, the following cloning steps were performed. The purified 1.2 kbp *XbaI/EcoRI* fragment carrying the 5′ portion of the coding region of U3 RNA (see Fig. 1A for restriction map) was ligated with the *XbaI/BamHI*-cleaved pBluescribe M13(+) vector. Following end-filling of the protruding *EcoRI* and *BamHI* sites with the Klenow fragment of DNA polymerase I, the plasmid was recircularized. The resulting new *XbaI/EcoRI* fragment (25 nt longer than the original one) was recovered and joined with the isolated 3.6 kbp *EcoRI/BamHI* fragment carrying the 3′ portion of the coding region of the U3 gene (see Fig. 1A), and inserted into the *XbaI/BamHI* sites of pBluescribe M13 (−).

Construction of 5′ promoter deletion mutants of the U3 gene was carried out by unidirectional exonuclease III digestion according to (37). The 5′ and 3′ protruding ends required for this step were generated by *XbaI* and *SphI*, respectively. The deletion mutants were preselected on the basis of their electrophoretic mobility on agarose gel and further characterized by sequencing (not shown). The resulting constructs (Fig. 7C) were cut out with *HindIII/BglII*, reinserted into the *HindIII/BglII* sites of the pGA492 binary vector (45) and transferred into *A. tumefaciens* A281 by direct DNA transfer (46). These bacterial cells were used to inoculate sterile *Nicotiana tabacum* SR1 leaf discs by the cocultivation method [cf. (45)]. Transformed callus tissue was selected by its resistance to kanamycin at 200 μg/ml. Shoots were regenerated from transformed calli as described by An (45). The kanamycin-positive plants were transferred to soil and grown in the greenhouse. Isolation and fractionation of nuclear RNA, conditions of molecular hybridizations as well as *in vitro* transcription of Probes A and B used in this experiment (Fig. 7A and B) are described above.

## RESULTS

### Isolation and organization of a tomato genomic locus for U3 RNA: Invariant structure of nine clones

To isolate tomato U3 genes, approximately $2\times10^5$ recombinant lambda phage plaques were screened with 3′ end-labeled tomato U3 RNA and nine positive clones were identified. *EcoRI* mapping of these phage DNAs revealed that all of these positive clones carried the same fragment of the tomato genome. This was confirmed by additional digestions with different restriction enzymes (some of which are entered in the map of the tomato U3 locus in Fig. 1A). Appropriate restriction fragments were subcloned into pBluescribe M13 vectors and their nucleotide sequences were determined.

To understand the genomic organization of U3 genes in the tomato genome we performed Southern blot analyses of tomato nuclear DNA. The results are shown in Fig. 1B. Complete digestion of the tomato DNA with *BglII* or *EcoRI* gave rise to one (3.75 kbp) and two (11 and 1.9 kbp) hybridizing bands, respectively. These results are in full agreement with the

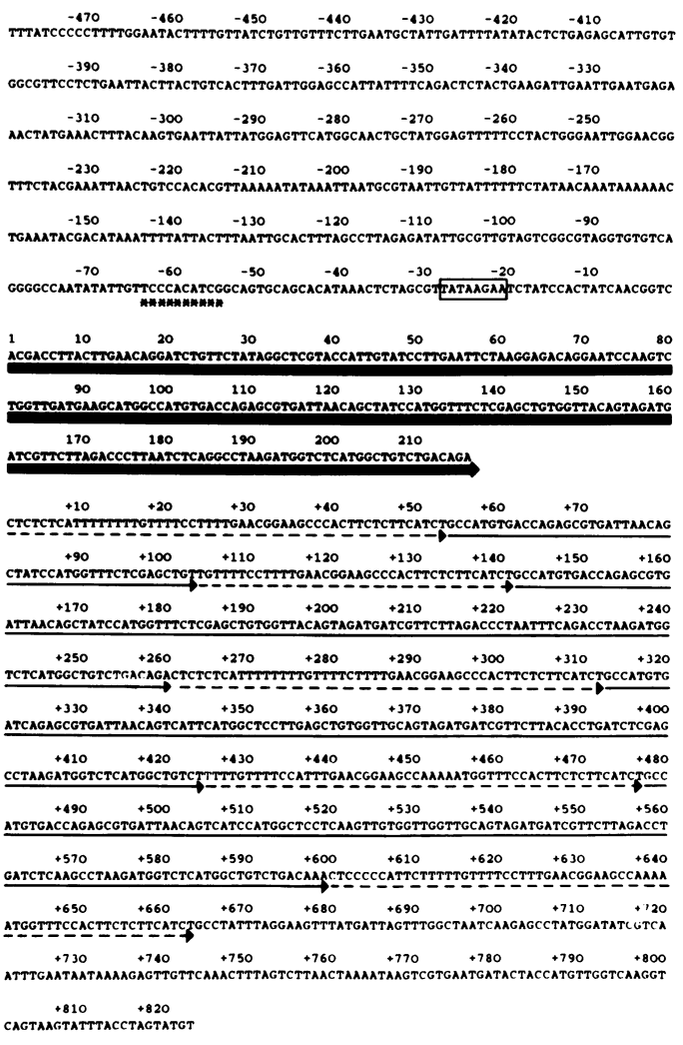

**Fig. 2.** Nucleotide sequence of the tomato U3 gene. The sequence of the non-coding DNA strand corresponding to the U3 RNA (positions 1 through 216) is marked by a bold arrow. Nucleotide positions in the 5′ and 3′ flanking regions are given negative (−) and positive (+) numbers, respectively. Upstream sequence elements conserved in all of the known plant U snRNA genes are indicated: the TATA-like element is boxed and the USE is denoted by asterisks below the sequence. The U3 pseudogenes and the direct repeats in the 3′-flanking region are indicated by fine arrows and dashed arrows, respectively.

restriction map of our U3 clone (see map in Fig. 1A and note that *EcoR*I splits the U3 coding region). Digestion with *EcoR*I and *Bgl*II yielded only the fragments which were predicted from the restriction map of the U3 gene, and digestion with various enzymes (*BamH*I, *Hind*III, *Kpn*I, *SpH*I, *Xba*I) having no restriction sites in the coding region gave rise to single hybridizing bands (not shown).

These results indicate that the isolated DNA fragment contained a U3 RNA-coding gene and that all U3-related sequences, including coding genes and pseudogenes, are carried by this fragment. However, our results do not exclude the possibility that U3-related sequences are present in some other fragment(s) that were not discernible under our experimental conditions.

### Identification and structural analysis of the coding region of the tomato U3 gene: Structural agreement with its transcript

Fig. 2 presents 480 nt of the 5′-flanking DNA sequence, 216 nt of the U3 coding region, and 823 nt of the 3′-flanking sequence determined according to the strategy outlined at the top of Fig. 1A.

The 5′ end of tomato U3 RNA was defined by primer extension assays (Fig. 3A) using a primer complementary to the non-coding (RNA-like) strand of the U3 gene from nucleotide 39 to 57. The gel electrophoretic mobility of the run-off transcript of AMV reverse transcriptase (Lane R) in relation to those of the DNA sequencing products (Lanes G, A, T and C) revealed that the reverse transcript terminates with a single band at the T residue underlined in the sequence 5′-GGTCG*T*GACCG-3′. This indicates that the 5′ terminus of U3 RNA commences with the complementary 5′-ACGACC-3′.

The 3′ terminus of U3 RNA was determined by direct RNA sequencing using base specific nucleases (Fig. 3B). This partial RNA sequence (64 nt) which corresponds to the 3′ portion of tomato U3 RNA (residues 153 through 216, see Fig. 2) and the DNA sequence of our U3 clone over this region are in perfect agreement.

### Primary and possible secondary structure of tomato U3 RNA: Failure to detect the canonical $m_3G$ cap structure, presence of four, phylogenetically highly conserved sequence blocks and peculiar secondary structure at the 5′ end

On the basis of the above experimental results the coding region of the tomato U3 gene is 216 nt long (Fig. 2). This conclusion was supported by the gel electrophoretic mobility of U3 RNA in denaturing polyacrylamide gels (data not shown). Since no 5′ end labeling was obtained after treatment of purified tomato U3 RNA with alkaline phosphatase, followed by incubation with polynucleotide kinase in the presence of [$\gamma$-$^{32}$P] ATP (not shown), we concluded that tomato U3 RNA has a 5′-terminal cap structure. We investigated the reactivity of rabbit anti-$m_3G$ IgG with purified tomato U3 RNA by an immune precipitation assay (Fig. 4). Anti-$m_3G$ antibodies selectively reacted with broad bean U2 RNA, which has a canonical $m_3G$ cap structure (11), but did not react with broad bean 5S RNA (negative control) and did not precipitate tomato U3 RNA.

To establish the possible secondary structure of tomato U3 RNA, we used an approach (see Materials and Methods) which yielded, in the case of S. pombe, S. cerevisiae, and rat U3 RNAs (not shown), secondary structure models that were basically the same as those already reported (16,18,21). Fig. 5 shows the possible secondary structure of tomato U3 RNA with the phylogenetically conserved regions identified. The most conspicuous feature is that the 5′ end region of tomato U3 RNA, in contrast to the U3 RNA molecules characterized to date from all higher eukaryotes, can be folded into two major stem-loops. Apart from this apparent difference, the secondary structure of tomato U3 RNA conforms to that of the other higher eukaryotic U3 RNAs [cf. (19,21)], including the occurrence of a long-range interaction (in tomato U3 RNA this is between nucleotides 88 to 108 and 194 to 214), as well as the presence of two stem-loops between these interacting regions.

### Structural analysis of the 5′ flanking region of the tomato U3 gene: Presence of both plant UsnRNA-specific transcription signals at an unusual spacing

The nucleotide sequence of the 5′ flanking region of the tomato U3 gene was compared with those of other plant spliceosomal UsnRNA genes or gene candidates sequenced to date (24−30). The promoter region of the tomato U3 gene carries only two sequence elements which are conserved and are present in the

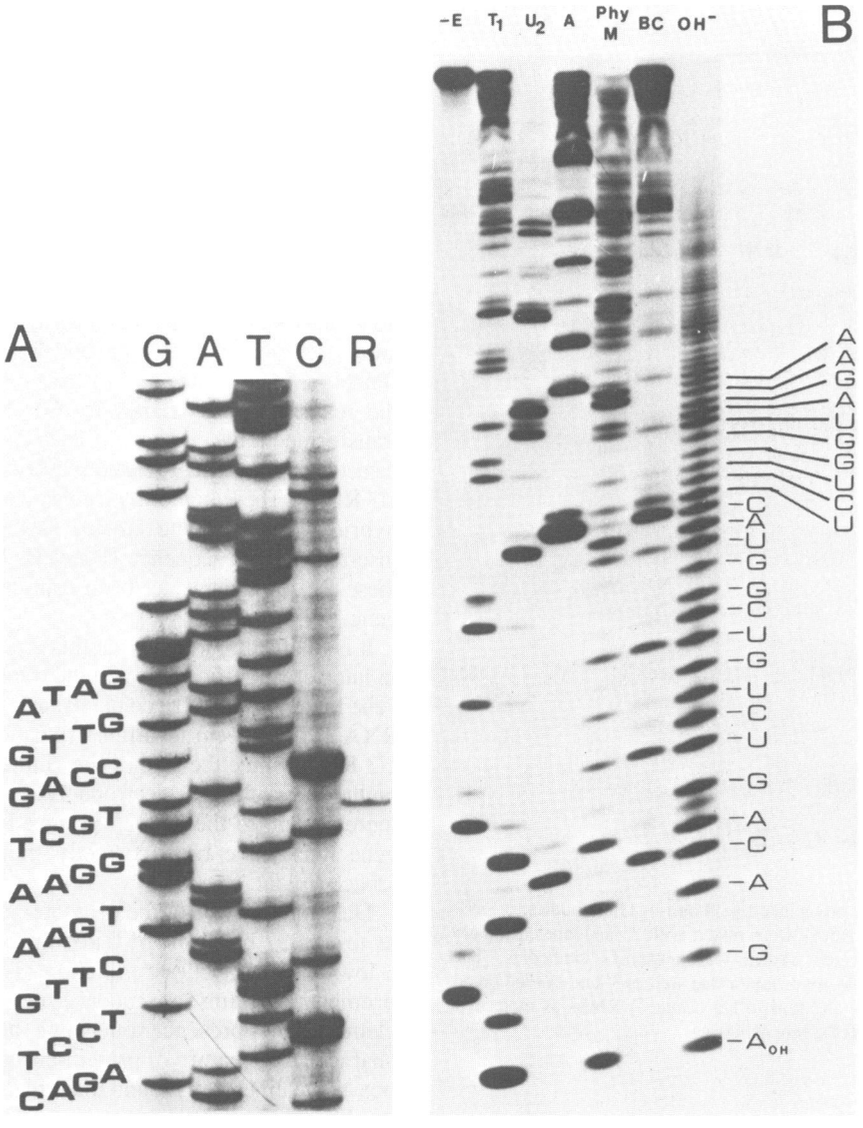


**Fig. 3.** Identification of the coding region of the tomato U3 gene. (**A**) Determination of the 5′ end of tomato U3 RNA by primer extension (for details see Materials and Methods). The sequencing reactions of the single-stranded recombinant pBluescribe plasmid harboring the coding strand of the tomato U3 gene are shown in *Lanes G, A, T* and *C*. The run-off product of reverse transcription is presented in *Lane R*. (**B**) Determination of the nucleotide sequence of the 3′ end of tomato U3 RNA. Autoradiograph of a 20% sequencing gel. Purified, 3′ end-labeled tomato U3 RNA was partially digested with RNase T1 (*Lane T1*, G-specific), RNase U2 (*Lane U2*, A-specific), RNase A (*Lane A*, C- and U-specific when they are 5′ to A), RNase Phy M (*Lane Phy M*, A- and U-specific) or RNase B. cereus (*Lane BC*, C- and U-specific). The autoradiographs obtained after partial alkaline hydrolysis (*Lane OH*$^-$) and without enzyme treatment (*Lane -E*) are also shown.

regulatory regions of all known plant UsnRNA genes or gene candidates, and also shown to be important for transcription of *Arabidopsis* U2 gene (31).

One of these is a TATA homology (5′-TATAAGAA-3′) at position −28 to −21 which is boxed in Fig. 2. The other one is a specific sequence block (5′-TCCCACATCG-3′), the so-called upstream sequence element (USE), from position −64 to position −55.

It is noteworthy that whereas the distance between these two promoter elements as well as between the TATA element and the cap site is conserved (32−36 and 22−27 nt, respectively) in the 5′ regulatory sequences all plant U1, U2 and U5 genes and gene candidates sequenced so far (24−30), in the tomato U3 gene the TATA-like element and the USE are spaced considerably closer to each other (26 nt) and also the TATA element is positioned only 20 nt upstream of the cap site.

### Structural analysis of the 3′ flanking region of the tomato U3 gene: Absence of the plant UsnRNA-specific putative transcription termination signal and presence of tandemly repeated U3 pseudogenes

The coding region of tomato U3 RNA is followed by a 27-nt-long pyrimidine-rich stretch (mostly T residues) with only two purine-insertions. This sequence is different from, and does not conform to, the consensus sequence of the putative termination signals of the plant U1, U2 and U5 genes and gene candidates (24−30).

A thorough sequence analysis of the non-coding region

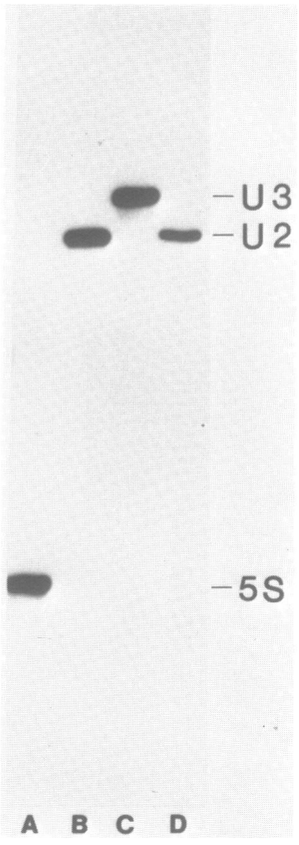


**Fig. 4.** Tomato U3 RNA is not precipitated with anti-$m_3G$ antibodies. 3′ end-labeled broad bean U2 and 5S RNAs were mixed with 3′ end-labeled tomato U3 RNA (20,000 cpm each) and treated with anti-$m_3G$ antibodies. The precipitable RNAs (*Lane D*) of the above mixture as well as 3′ end-labeled broad bean 5S (*Lane A*), U2 (*Lane B*) and tomato U3 (*Lane C*) RNAs as molecular markers were separated on a 10% sequencing gel.

downstream of the U3 coding gene revealed that the 3′ flanking sequence between positions +1 and +53 (Fig. 2) is subsequently repeated four times in slightly altered forms. Interspersed among these elements are four tandemly repeated DNA segments (pseudogenes) that show substantial similarity among themselves and with the coding region of the U3 gene. Fig. 6A represents the schematic structure of the tomato U3 locus with the 3′ non-coding region carrying the four pseudogenes (U3.1ps to U3.4ps) for U3 RNA.

The most significant feature of the four U3 pseudogenes is that they are truncated at their 5′ ends, and that, surprisingly, all of them start precisely with the nucleotide sequence 5′-GCCA...-3′ corresponding to the sequence of the U3 coding gene from position G97 (Fig. 2). In addition, two of them (U3.1ps and U3.3ps) are truncated at their 3′ ends as well. To systemize the individual copies of the repeated DNA fragments, the four U3 pseudogenes (U3.1ps to U3.4ps) were compared with the corresponding region of the U3 coding gene (Fig. 6B) and the 3′ direct repeats flanking the pseudogenes (R2 to R5) with the adjacent 3′ flanking sequence of the U3 gene (R1) (Fig. 6C). Disregarding the deletions at the 3′ end regions of U3.1ps and U3.3ps, the nucleotide sequences of U3.1ps, U3.2ps, U3.3ps and U3.4ps revealed 100, 97.6, 87.7 and 75.6% sequence similarity to the U3 coding region, respectively.

### Expression of the tomato U3 gene in transgenic tobacco plants: The USE is needed for efficient gene expression

To check the transcriptional activity of our U3 gene in transgenic tobacco plants and to test the functional significance of the two conserved sequence blocks of the promoter region, tagged and 5′-deleted derivatives of the tomato U3 gene were constructed (Fig. 7C) and used for transformation of tobacco cells.

RNA extracted from the nuclear fractions of transformed plants was analyzed by Northern blotting (Fig. 7A and B, Lanes C−G). Purified tomato U3 RNA (Lanes A) and total nuclear RNA from non-transformed tobacco (Lanes B) were run as controls. The antisense tomato U3 RNA probe hybridized to tomato U3 RNA (Panel A, Lane A) and to the tobacco U3 RNAs transcribed from the resident genes (Lanes B−G). Moreover, in the case of transformed tobacco plants (Lanes C−G) the U3-specific probe detected additional U3-related sequences migrating above the host U3 RNAs. These higher-molecular-weight RNAs showed specific hybridization with the labeled ssDNA complementary to the inserted marker sequence (Panel B, Lanes C−G) proving that these RNAs had in fact been transcribed from the transferred gene.

It is interesting to note that this type of tagging did not appear to interfere significantly with the transcription of the tomato U3 gene in transgenic tobacco plants, and the stability of the chimeric RNA did not seem to differ dramatically from that of tobacco U3 RNA. Since the size of the chimeric RNA is in agreement, on the basis of its gel electrophoretic mobility, with the expected increase due to the tag (25 nt), we conclude that the chimeric gene must have been correctly transcribed in the transgenic tobacco plant.

Deletion analysis of the promoter of the U3 gene showed that as few as 37 bp of the 5′ flanking sequence were sufficient for a low level of accurate initiation. This region of the U3 gene promoter contains a sequence motif resembling the TATA element, the presence of which at a similar position is a characteristic feature of protein-encoding genes and this TATA box-like element was found in all plant UsnRNA genes and gene candidates (24−30) analyzed so far. These results suggest that TATA homologies of plant UsnRNA genes are functionally important and are responsible for the accuracy but not for the efficiency of transcription initiation. Progressive deletions from position −1200 to −135 (Lanes C−F) did not considerably affect efficient U3 transcription, but deletion to −37 reduced the accumulation of U3 RNA 30- to 40-fold as compared to the original activity (Lanes C and G). These results indicate that sequences between positions −135 and −37 contain an element capable of activating transcription of the tomato U3 gene in transgenic tobacco plants. This region spans the highly conserved USE of the tomato U3 gene that may thus function as a plant UsnRNA-specific promoter element.

## DISCUSSION

The results presented in this paper yield, first of all, phylogenetic support to the notion that U3 RNA is set apart from the spliceosomal UsnRNAs with respect to both structure and expression of its gene, and show that in plants this divergence is characterized by some particular features that are different from those of metazoan and yeast U3 RNAs.

Whereas the evolutionarily highly conserved Boxes A to D, and the long-range interaction domain are both present in tomato

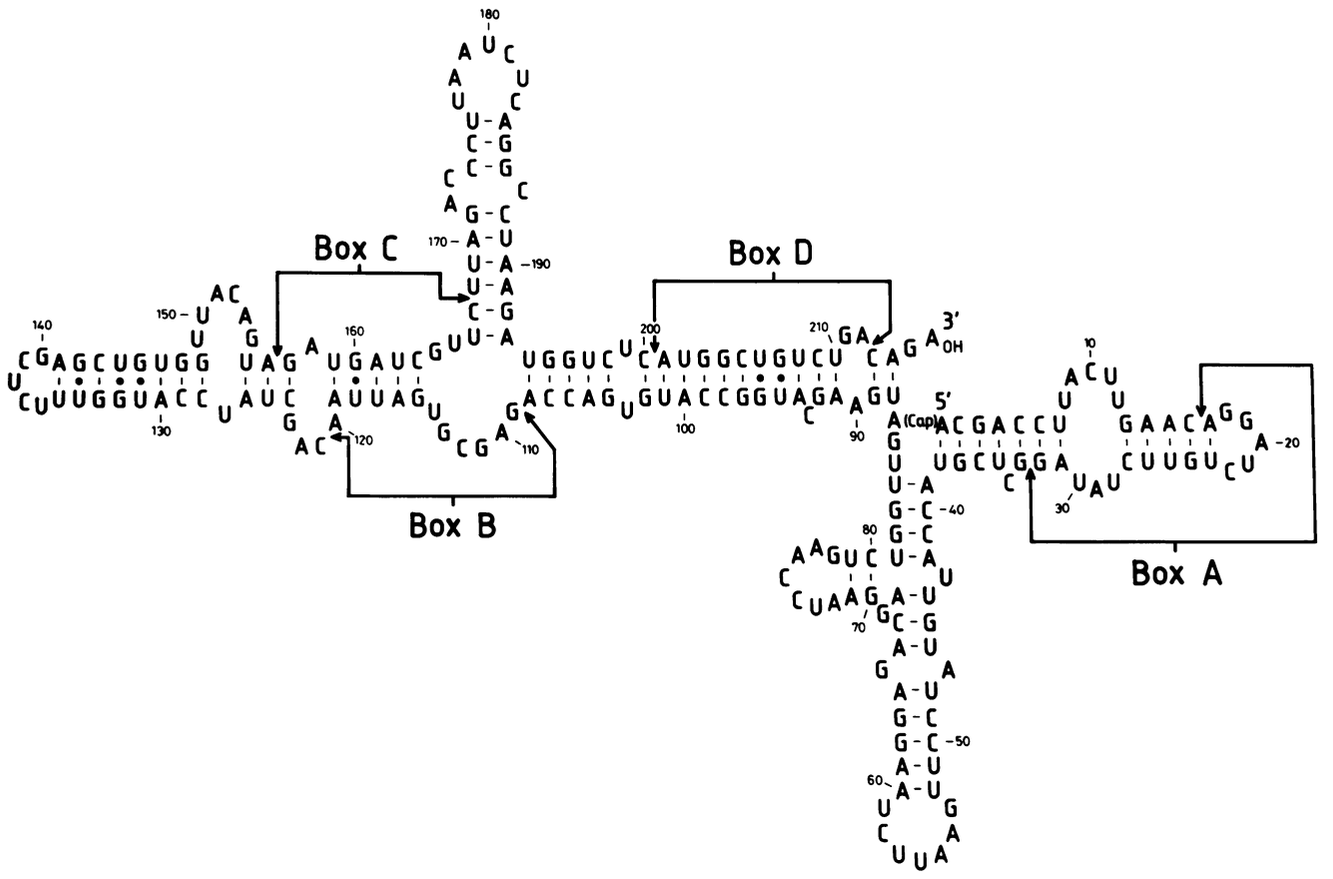


**Fig. 5.** Proposed secondary structure of tomato U3 RNA. The phylogenetically highly conserved Boxes A to D are indicated.

U3 RNA (See Fig. 5 and Results), we could not detect the 5′ end canonical cap structure in this molecule (Fig. 4). Although the possibility that the $m_3G$ cap in the tomato U3 RNA is not available for reaction with specific antibodies under the experimental conditions we used cannot be ruled out, our findings rather suggest (see Results) that tomato U3 RNA has an unusual cap structure. In a number of experiments in our laboratory with deproteinized tomato nuclear RNAs or with purified broad been U3 RNA we have never observed precipitation of plant U3 RNA with anticap antibodies (data not shown). Krol et al. (10) also failed to precipitate pea U3 RNA using antibodies directed against the $m_3G$ cap structure. The possible presence of two major stem-loops in the 5′ half of tomato U3 RNA casts some doubt on the validity of the suggestion (16) that this type of secondary structure would be characteristic of lower eukaryotic U3 RNAs as opposed to higher eukaryotic U3 RNAs the 5′ end portions of which would fold into a single stem-loop structure. We favour the alternative view, also proposed by Porter et al. (16), that the 5′ end portion of U3 RNAs may have no secondary structure at all. This assumption seems to be supported by the fact that the difference in free energy between the two alternative 5′ end secondary structures of S. cerevisiae U3 RNA is negligible. An analysis of the possible impact of the phylogenetically conserved structural elements of U3 RNA on its alleged function in pre-rRNA processing will be published elsewhere.

As described in the Introduction, in the regulation of the transcription by Pol. II of the genes for the nucleolar U3 RNA in mammals a so-called 'U3 Box' may be involved, in addition to the PSE and DSE. The presence of this cis-acting element may ensure that transcription of the U3 gene be independent from that of the genes coding for the spliceosomal U1, U2, U4 and U5 RNAs. A major conclusion from our work described in this paper is that in plants this type of alleged 'uncoupling' of gene expression may occur by a basically different mechanism. In the tomato *bona fide* U3 gene we found the same upstream transcription signals as those present in the plant spliceosomal UsnRNA genes, their spacing, however, differed from that of the latter. In this respect is resembled a tomato U6 gene candidate (47) and also, more importantly, an *Arabidopsis bona fide* U6 gene (F. Waibel and W. Filipowicz, personal communication) in both of which the spacing between the TATA element and the USE as well as the distance of these promoter elements from the cap site is reduced as compared to that in the plant genes for U1, U2 and U5 RNAs. In vertebrates, the U6 gene is transcribed by Pol. III (48,49). If we assume that the same is true of plant systems, it is tempting to speculate that in plants the genes for U3 RNA are transcribed by a Pol. III-type enzyme rather than by Pol. II. This hypothesis is supported by the following additional data: (i) In the 3′ flanking region of both the tomato U3 gene and the tomato U6 gene candidate (47) the '3′ consensus sequence', a possible transcription termination signal characteristic of the plant U1, U2 and U5 genes (24−30) is absent. (ii) Nine nt residues downstream of the 3′ end of the coding sequence there is an eight-nt-long pyrimidine block flanked by purines in both the U3 gene [$(T)_8$] and the U6 gene candidate [(TTCTTTTC)]. This could be a transcription termination signal

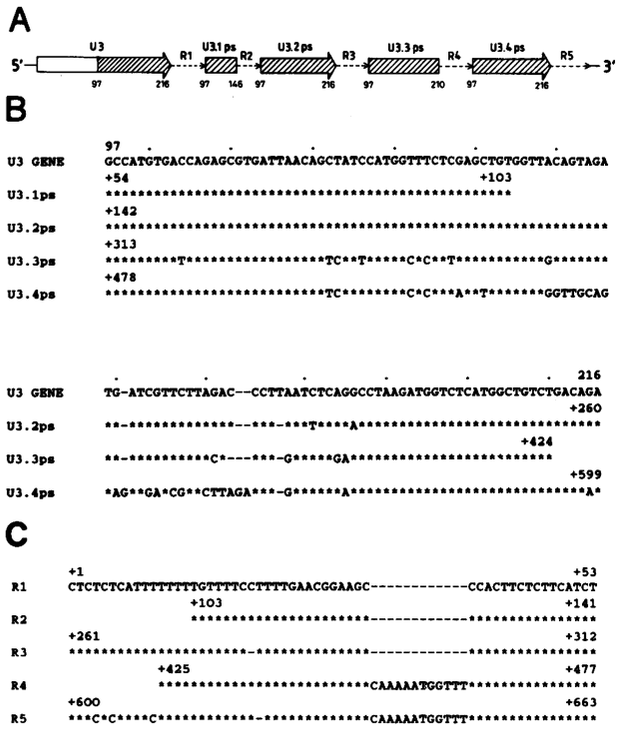


**Fig. 6.** Organization of pseudogenes in the 3′ flanking region of the tomato U3 gene. (**A**) Schematic structure of the tomato U3 locus. The arrows and the boxes indicate the U3 RNA-specific sequences. The 3′ end portion of the U3 coding region (U3) and its repeating pseudogene versions (U3.1ps to U3.4ps) are cross-hatched. Numbering below the U3 pseudogenes corresponds to that of the coding region of the U3 gene (on the left). The imperfect direct repeat sequences (R1 to R5) are indicated by dashed arrows. (**B**) and (**C**) Comparison of the repeating sequences in the 3′ flanking region of the tomato U3 gene. Asterisks stand for identical nucleotides and dashes for deletions. Mismatched bases and sequence insertions are identified. The numbers at the beginning and at the end of each sequence correspond to the numbering in Fig. 2. (**B**) The nucleotide sequences of the tandemly repeated U3 pseudogenes (U3.1ps to U3.4ps) are compared with the sequence of the U3 coding gene. (**C**) The nucleotide sequences of the imperfect direct repeats (R1 to R5) flanking the U3 pseudogenes are compared with each other.

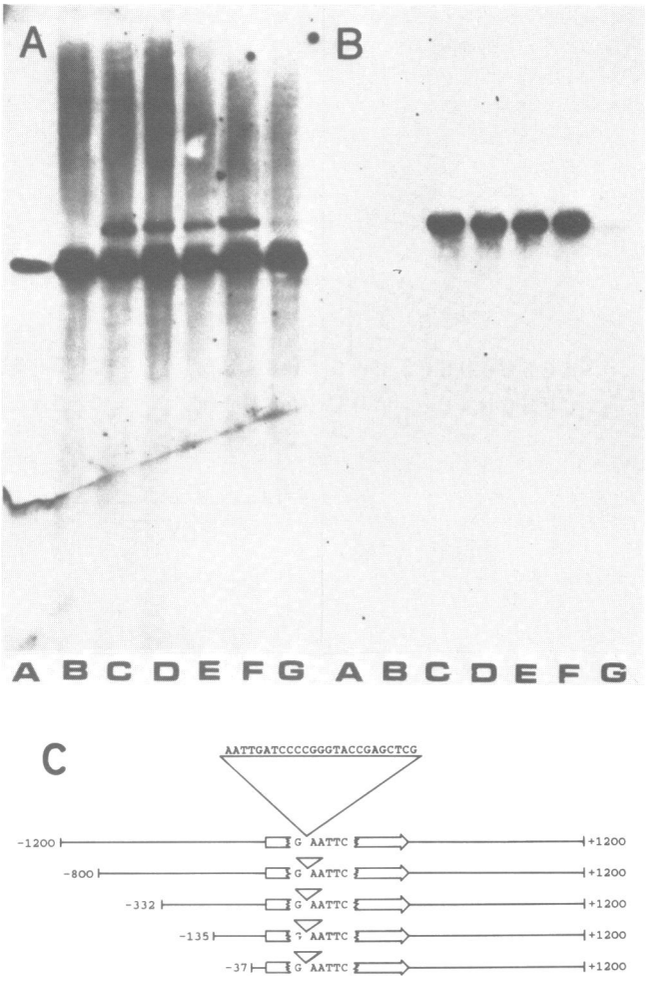


**Fig. 7.** Expression of the tomato U3 gene in transgenic tobacco plants. (**A**) and (**B**) About 10 ng purified tomato U3 RNA (*Lanes A*), 4 μg tobacco nuclear RNA (*Lanes B*), and 4 μg nuclear RNA from transgenic tobacco plants transformed with constructs (**C**) −1200 (*Lanes C*), −800 (*Lanes D*), −332 (*Lanes E*), −135 (*Lanes F*) and −37 (*Lanes G*) were fractionated on a 10% denaturing polyacrylamide gel. The RNAs were transferred to nitrocellulose filter and probed with (**A**) *in vitro* synthesized antisense tomato U3 RNA (Probe A, see Materials and Methods) or with (**B**) ssDNA complementary to the inserted marker sequence (Probe B, see Materials and Methods). (**C**) Schematic diagram showing the structures of insertion mutants obtained by progressive 5′ deletions of the tomato U3 gene. The 25-nt insertion at the internal *EcoRI* site was derived from the polylinker sequence of the pBluescribe M13 vector. The U3 coding regions are indicated by arrows. Numbers at the beginning of the deletion constructs indicate positions relative to the cap site of mature U3 RNA and correspond to the numbering in Fig. 2.

for a Pol. III-type enzyme. (iii) Neither plant U6 RNA nor plant U3 RNA could be precipitated by anti-$m_3G$ antibodies. U6 RNA from all organisms studied so far is known to lack the canonical $m_3G$ cap structure at its 5′ terminus (1) and to possess, instead, a γ-monomethyl phosphate cap structure (50).

In the genome of a number of metazoan species UsnRNA-related sequences were found to be present in multiple copies. The majority of these loci proved to be pseudogenes rather than *bona fide* genes [cf. (22)]. Metazoan pseudogenes fall into two main groups: those generated by RNA-mediated mechanisms and those arisen through DNA-mediated events [cf. (51,52)]. Strangely enough, searches in plant genomes for UsnRNA-related sequences did not reveal the presence of pseudogenes either for U2 and U5 RNAs in *Arabidopsis thaliana* (27,30) or for U1 RNA in either common bean (24) or soybean (25).

In the tomato genome, however, we found a U1 pseudogene probably generated by RNA-mediated mechanisms (53). In this paper we report the isolation of U3 pseudogenes which are present in the vicinity of a *bona fide* U3 gene. This particular location of the U3 pseudogenes apparently generated by DNA-mediated mechanisms allowed us to deduce the chronology of events leading to the emergence of the tandemly repeated U3 pseudogenes (Fig. 6A). As a first step, the U3.4ps could have been inserted along with its adjacent terminator sequence by the mechanism of unequal recombination. A possible fossil of this misalignment might be the sequence motif adjacent to the 3′ end of the R5 repeat (5′-GCCTATTT-3′; positions +664 to +671) which shows sequence similarity to the starting sequence motives of U3 pseudogenes (5′-GCCATGT-3′). Next, the U3.3ps could have been inserted by unequal crossing-over between the coding region of the U3 gene and the U3.4 pseudogene. Subsequently, U3.2ps and U3.1ps might have arisen by similar mechanisms. The 12-nt segment in the repeating R4 and R5 sequences (Fig. 6C) should have been deleted after the insertion of U3.3ps as this shortened repeated flanking sequence (R3) must be the progenitor of all subsequent repeated 3′ flanking sequences (R1 and R2). Deletions at the 3′ regions of U3.1ps and U3.3ps which

overlap with the 5′ end regions of R2 and R4 as well, might have been secondary events and their chronological order cannot be determined with precision.

This is the first report (i) on the complete nucleotide sequence of a plant U3 RNA, (ii) on the functional analysis of a plant U3 RNA *bona fide* gene, (iii) on the existence of plant UsnRNA pseudogenes generated by DNA-mediated mechanisms and (iv) on a genomic locus in which tandemly repeated UsnRNA pseudogenes are located in the immediate vicinity of a *bona fide* UsnRNA gene, allowing thereby the construction of a chronological order of events leading to their generation.

## ACKNOWLEDGEMENTS

Thanks are due to Dr. R.W. Breidenbach for kindly supplying us with a tomato genomic library, to Professor R. Lührmann for his kind gift of anti-$m_3G$ IgG, to Miss H. Hegyi for establishing the possible secondary structure of the tomato U3 RNA, to Drs W. Filipowicz and F. Waibel for helpful dicussions and a critical review of the manuscript, to Mr. B. Dusha for preparing the photographs, to Ms. G. Nöllenburg for drawing the figures and, last but not least, to Ms. M. Csécsei for typing the manuscript. This work was supported by Grants OKKFT(Tt)/1986 and OTKA 174/1988 to F.S. as well as by a grant to F.S. from the Hungarian Ministry of Industry.

## REFERENCES

1. Reddy,R. and Busch,H. (1988) In Birnstiel,M.L. (ed.), Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, Springer Verlag, Berlin, pp. 1–31.
2. Guthrie,C. and Patterson,B. (1988) *Annu. Rev. Genet.*, **22**, 387–419.
3. Steitz,J.A., Black,D.A., Gerke,V., Parker,K.A., Kramer,A., Frendewey,D. and Keller,W. (1988) In Birnstiel,M.L. (ed.) Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Springer Verlag, Berlin, pp. 115–154.
4. Prestayko,A.W., Tonato,M. and Busch,H. (1970) *J. Mol. Biol.*, **47**, 505–515.
5. Calvet,J.P. and Pederson,T. (1981) *Cell*, **26**, 363–370.
6. Epstein,P., Reddy,R. and Busch,H. (1984) *Biochemistry*, **23**, 5421–5425.
7. Maser,R.L. and Calvet,J.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6523–6527.
8. Stroke,I.L. and Weiner,A.M. (1989) *J. Mol. Biol.*, **210**, 497–512.
9. Kass,S., Tyc,K., Steitz,J.A. and Sollner-Webb,B. (1990) *Cell*, in press.
10. Krol,A., Ebel,J., Rinke,J. and Lührmann,R. (1983) *Nucleic Acids Res.*, **11**, 8583–8594.
11. Kiss,T., Antal,M. and Solymosy,F. (1987) *Nucleic Acids Res.*, **15**, 1332.
12. Kiss,T., Antal,M. and Solymosy,F. (1987) *Nucleic Acids Res.*, **15**, 543–560. **16**, 2734.
14. Kiss,T., Jakab,G., Antal,M. Pálfi,Z., Hegyi,H., Kis,M. and Solymosy,F. (1988) *Nucleic Acids Res.*, **16**, 5407–5426.
15. Reddy,R. (1988) *Nucleic Acids Res.*, **16**, r71-r85.
16. Porter,G.L., Brennwald,P.J., Holm,K.A. and Wise,J.A. (1988) *Nucleic Acids Res.*, 16, 10131–10152.
17. Jeppesen,C., Stebbins-Boaz,B. and Gerbi,S.A. (1988) *Nucleic Acids Res.*, 16, 2127–2148.
18. Hughes,J.M.X., Konings,D.A.M. and Cesareni,G. (1987) *EMBO J.*, **6**, 2145–2155.
19. Parker,K.A. and Steitz,J.A. (1987) *Mol. Cell. Biol.*, 7, 2899–2913.
20. Kiss,T., Tóth,M. and Solymosy,F. (1985) *Eur. J. Biochem.*, 152, 259–266.
21. Konings,D.A.M. and Hogeweg,P. (1989) *J. Mol. Biol.*, **207**, 597–614.
22. Dahlberg,J.E. and Lund,E. (1988) In Birnstiel, M.L. (ed.), Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Springer Verlag, Berlin, pp. 38–70.
23. Parry,H.D., Scherly,D. and Mattaj,I.W. (1989) *Trends Biochem. Sci.*, **14**, 15–19.
24. van Santen,V.L. and Spritz,R.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 9094–9098.
25. van Santen,V.L., Swain,W. and Spritz,R.A. (1988) *Nucleic Acids Res.*, 16, 4176.
26. Abel,S., Kiss,T. and Solymosy,F. (1989) *Nucleic Acids Res.*, 17, 6319–6337.
27. Vankan,P. and Filipowicz,W. (1988) *EMBO J.*, 7, 791–799.
28. Hanley,B.A. and Schuler,M.A. (1989) *Nucleic Acids Res.*, **17**, 10106.
29. Brown,J.W.S. and Waugh,R. (1989) *Nucleic Acids Res.*, **17**, 8991–9001
30. Vankan,P. Edoh,D. and Filipowicz,W. (1988) *Nucleic Acids Res.*, 16, 10425–10440.
31. Vankan,P. and Filipowicz,W. (1989) *EMBO J.*, 8, 3875–3882.
32. Mazan,S. and Bachellerie,J.-P. (1988) *J. Biol. Chem.*, **263**, 19461–19467.
33. Yuan,Y. and Reddy,R. (1989) *Biochim. Biophys. Acta*, **1008**, 14–22.
34. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
35. Benton,W.D. and Davis,R.W. (1977) *Science*, 196, 180–182.
36. Helms,C., Graham,M.Y., Dutchik,J.E. and Olson,M.V. (1985) *DNA*, **4**, 39–49.
37. Henikoff,S. (1984) *Gene*, 28, 351–359.
38. Russel,M., Kidd,S. and Kelley,M.R. (1986) *Gene*, 45, 333–338.
39. Sanger,F., Nicklen,S. and Coulson,A.R.(1977) *Proc. Natl. Acad. Sci. USA*, 74, 5463–5467.
40. Melton,D.A., Krieg,P.A., Rebagliatti,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
41. Kiss,T., and Solymosy,F. (1987) *Acta Biochem. Biophys. Hung.*, 22, 1–5.
42. D'Alessio,J.M. (1982) In Rickwood,D. and Hames,B.D. (eds), Gel Electrophoresis of Nucleic Acids. IRL Press, Oxford, pp. 185–187.
43. Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl. Acad. Sci. USA*, 83, 9373–9377.
44. Zucker,M. and Stiegler,P. (1981) *Nucleic Acids Res.*, 9, 133–148.
45. An,G. (1986) *Plant Physiol.*, 81, 86–91.
46. An,G. (1987) *Methods Enzymol.*, 153, 292–305.
47. Szkukálek,A., Kiss,T. and Solymosy,F. (1990) *Nucleic Acids Res.*, **18** (in press).
48. Kunkel,G.R., Maser,R.L., Calvet,J.P. and Pederson,T. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8575–8579.
49. Reddy,R., Henning,D., Das,G., Harless,M. and Wright,D. (1987) *J. Biol. Chem.*, 262, 75–81.
50. Singh,R. and Reddy,R. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 8280–8283.
51. Denison,R.A. and Weiner,A.M. (1982) *Mol. Cell. Biol.*, 2, 815–828.
52. Hammarström,K., Bark,C., Westin,G., Zabielski,J. and Pettersson,U. (1986) *Chemica Scripta*, 12B, 165–170.
53. Kiss,T., Abel,S. and Solymosy,F. (1989) *Plant Mol. Biol.*, 12, 709–711.